



Нейросетевые подходы к рекомендациям и NMF

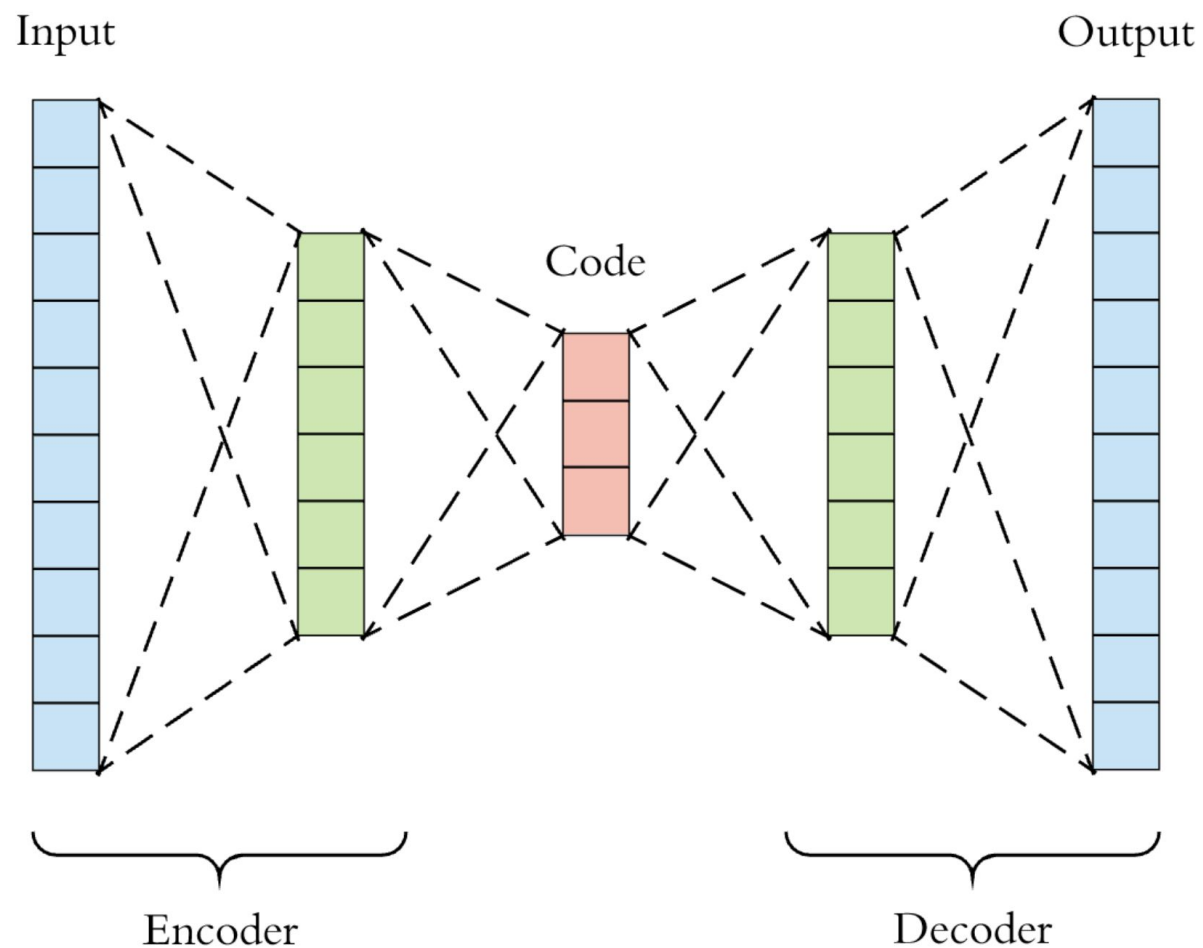


Волохов Глеб
ML разработчик команды MTC Kion

Нейросетевые подходы к рекомендациям

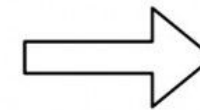
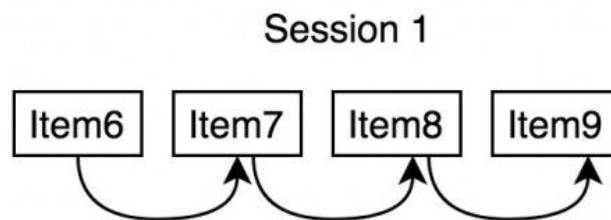
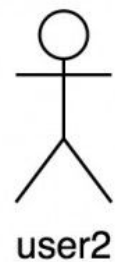
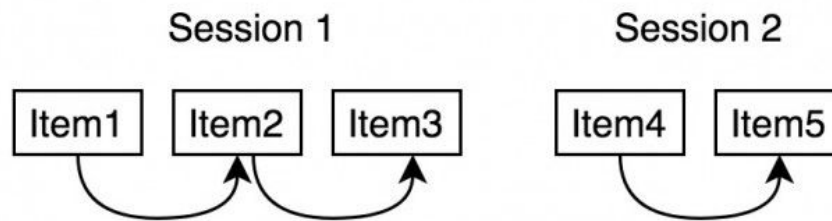
- Автоэнкодеры
- $i2v$
- NMF

Автоэнкодеры



Word2vec

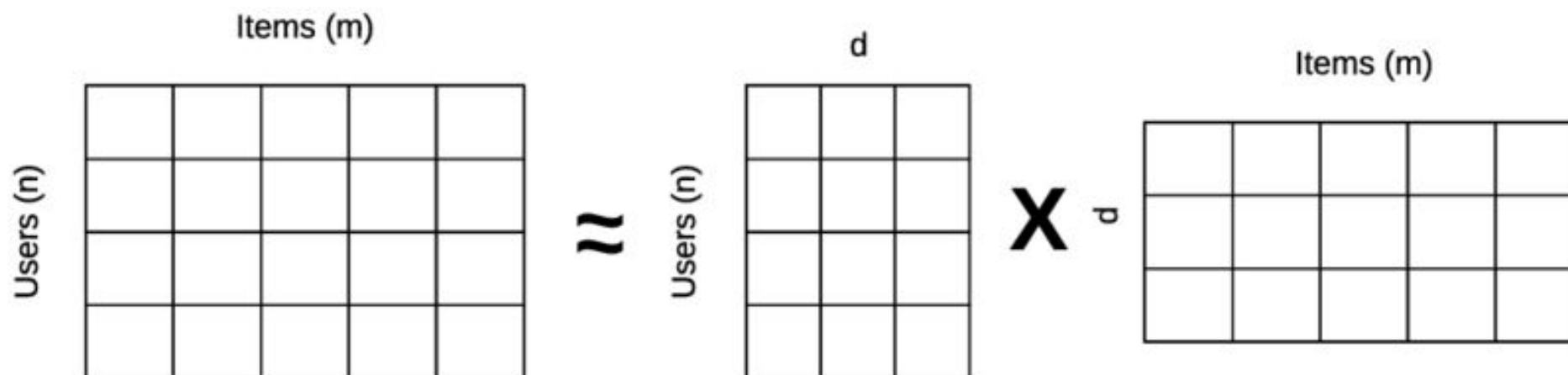
User's contacts



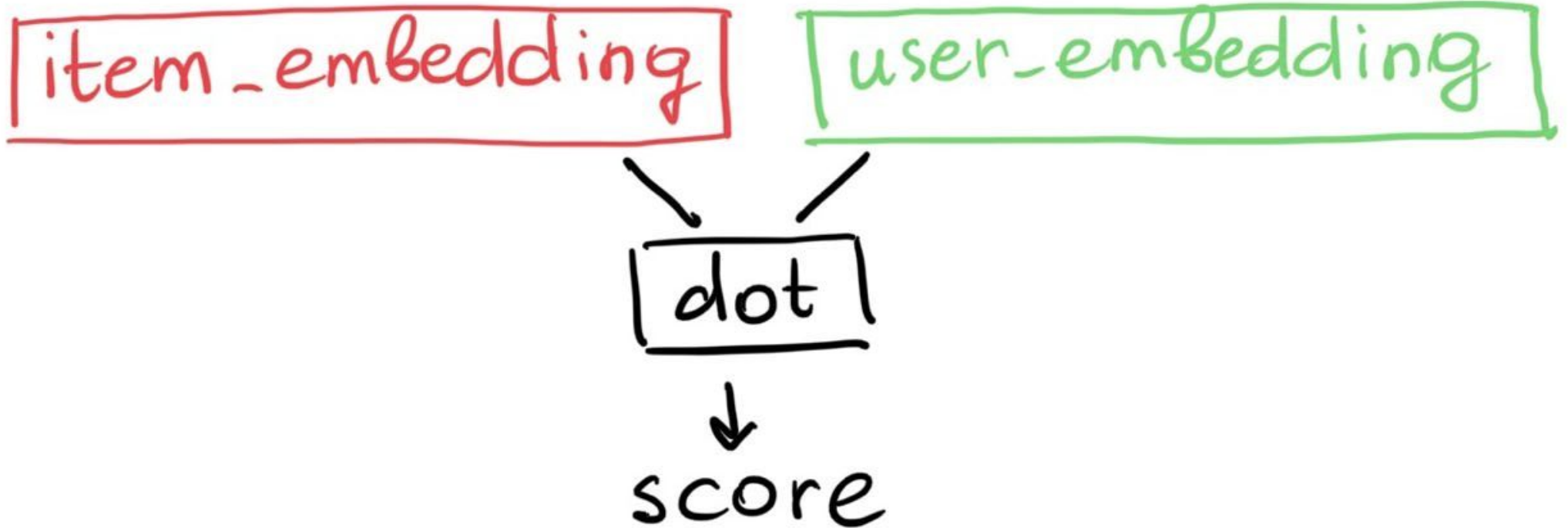
Dataset

(Item1, Item2)
(Item2, Item3)
(Item4, Item5)
(Item6, Item7)
(Item7, Item8)
(Item8, Item9)

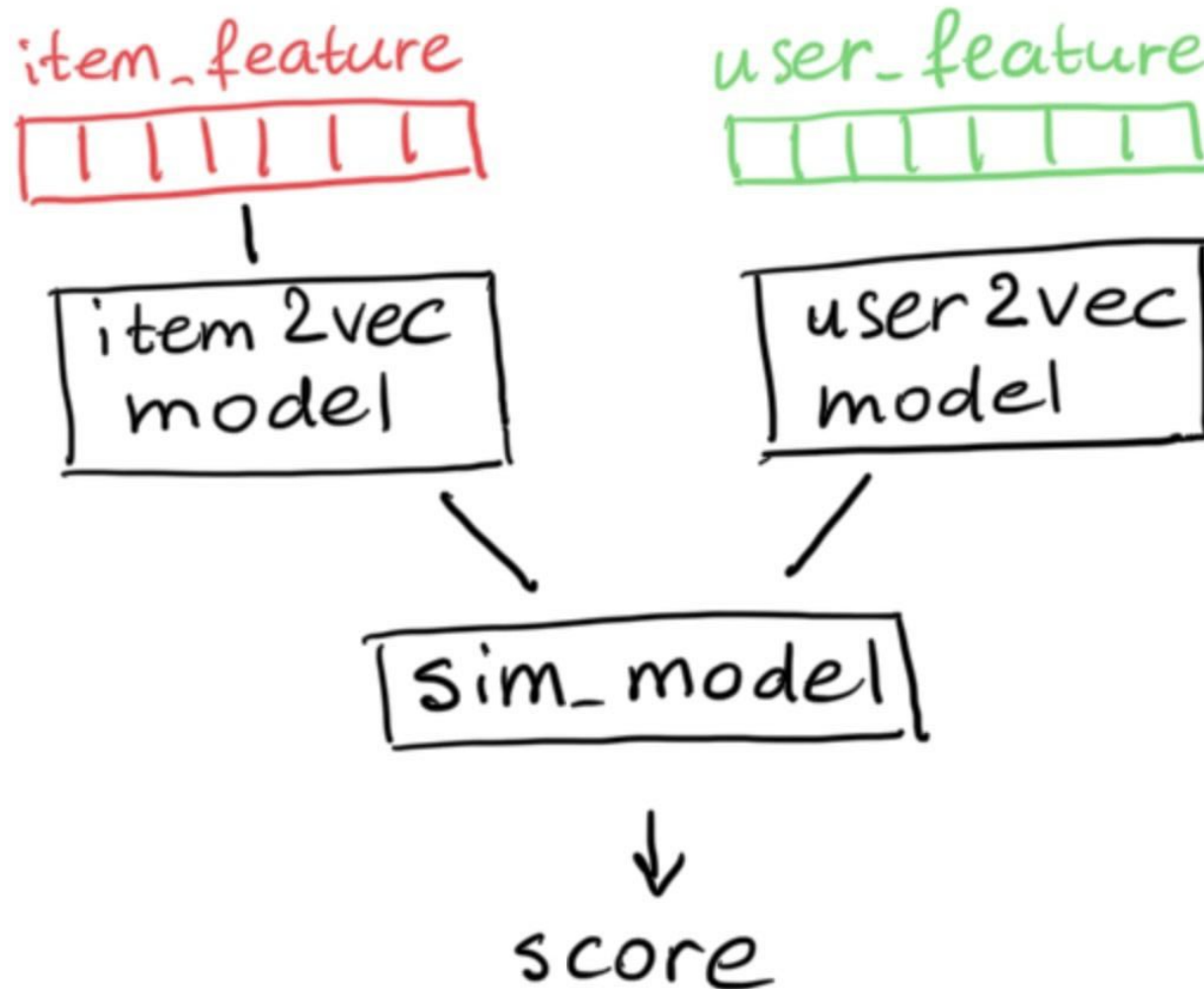
Матричная факторизация



MF (Matrix Factorization)



NMF (Neural Matrix Factorization)

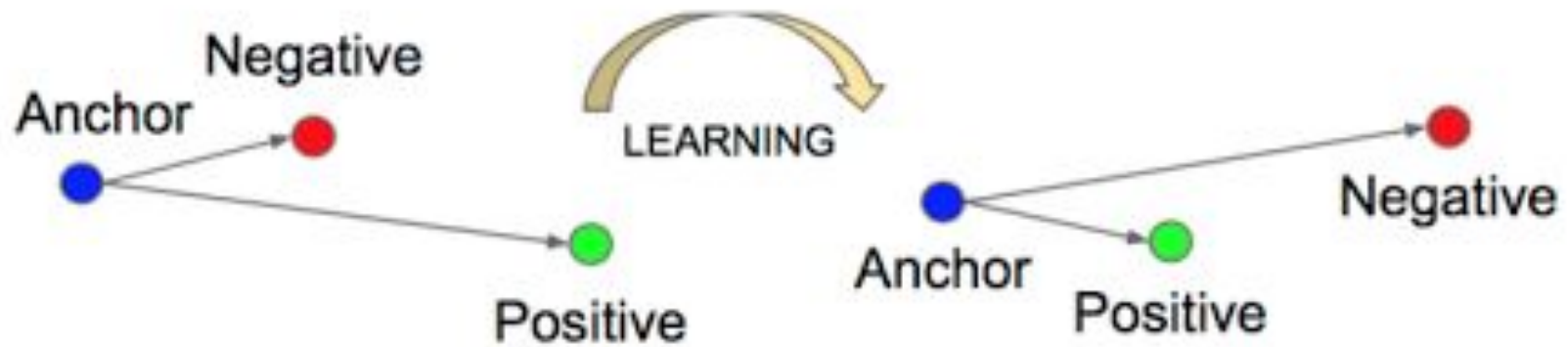




Что учим?

- *Классификация*
- *Регрессия*
- *learn2rank*

TripletLoss



$$pos_dist = ||user2vec(U) - item2vec(P)||^2$$

$$neg_dist = ||user2vec(U) - item2vec(N)||^2$$

$$basic_loss = pos_dist - neg_dist$$

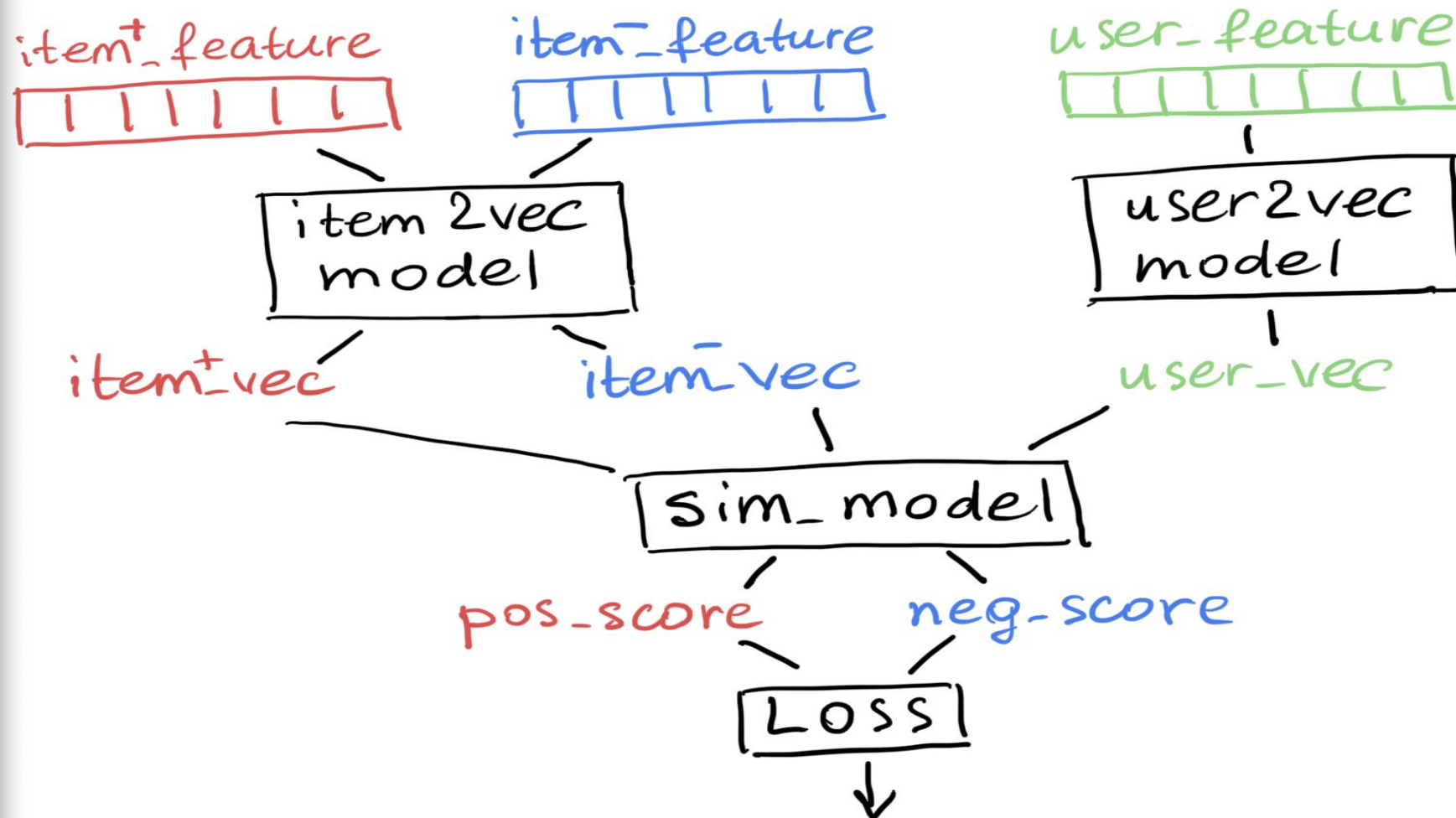
$$TPL = \max(basic_loss + \alpha, 0)$$

Еще немного l2r loss

$$L_{margin}(a, p, n) = \sum \max(0, \text{sim}(a, p) - \text{sim}(a, n) + \varepsilon)$$

$$L_{BPR}(a, p, n) = \sum (1 - \sigma(\text{sim}(a, p) - \text{sim}(a, n)))$$

Triplet loss + NMF \Rightarrow сиамские сети



Семплирование негатива

- *Случайные айтемы без взаимодействия*
 - *Negative mining*
 - *Иерархия отношений*
 - *Увидел, но не кликнул*
-
- *feedback loop*

Модель user2vec и фичи юзера



Холодные фичи

- Доступны до прихода
- Характеризуют группу пользователей

Горячие фичи

- Генерируются по мере пользования
- Характеризуют пользователя



Почему нейросети?

- *Нелинейность*
- *Гибкость в настройке*
- *Deer learning ®*
- *Широкие возможности работы с холодными элементами*
- *Моделирование последовательностей событий*
- *Cuda и GPU бесплатно*
- *Быстрый прод на векторах*



Почему не нейросети?

- *Слишком гибко*
- *Время разраб*
- *Нужно больше данных*
- *Интерпретация результатов*



Когда нейросети?

- *Много данных и разных фичей*
- *Есть базовая модель и не ясно как улучшать*
- *Быстрая ротация item*
- *Кросс рекомендации*
- *Нужно векторное представление пользователей и айтемов*



Мини советы

- *Не усложняй!*
- *Логируй эксперименты*
- *Архитектура не самое важное*
- *Думай над стратегиями генератора*
- *Думай над фичами*
- *Предобучай i2v модель*

Конец



Created by  **MTC**

Ссылки

[Item2vec от авито](#)

[ranking loss](#)

[Негативное семплирование](#)